**The K-means clustering algorithm**

The K-means clustering algorithm is a well-known and widely used clustering algorithm. It partitions the data into k clusters, k being a pre-determined parameter, which the user determines, as part of an educated guess based on prior knowledge of the data. Each data set belongs to one and only one partition in a non-overlapping manner.

The algorithm works in the following way:

* It specifies the number of clusters, k
* It initializes k ‘centroid’ data points, which act as the center for each of the clusters, by randomly selecting k data points from the data set.
* In an iterative step, the centroid data points are reshuffled, and then:
  + The sum of squares between all data points and the centroids are computed
  + Each data point is assigned to a cluster, represented by a centroid
  + The new centroid for a given cluster will be the average of all data points

**References:**

1. towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a